

# STATISTICAL EFFICIENCY OF STRUCTURED CPD ESTIMATION APPLIED TO WIENER-HAMMERSTEIN MODELING

José Henrique de M. Goulart<sup>\*,§</sup>, Maxime Boizard<sup>\*,†</sup>, Rémy Boyer<sup>\*\*</sup>, Gérard Favier<sup>\*</sup> and Pierre Comon<sup>‡</sup>

<sup>\*</sup> I3S, Université Nice Sophia Antipolis, CNRS, France

<sup>\*\*</sup> L2S, Université Paris-Sud, CNRS, CentraleSupélec, France

<sup>†</sup> SATIE, ENS Cachan, France

<sup>‡</sup> GIPSA-Lab, Université de Grenoble, CNRS, France

## ABSTRACT

The computation of a structured canonical polyadic decomposition (CPD) is useful to address several important modeling problems in real-world applications. In this paper, we consider the identification of a nonlinear system by means of a Wiener-Hammerstein model, assuming a high-order Volterra kernel of that system has been previously estimated. Such a kernel, viewed as a tensor, admits a CPD with banded circulant factors which comprise the model parameters. To estimate them, we formulate specialized estimators based on recently proposed algorithms for the computation of structured CPDs. Then, considering the presence of additive white Gaussian noise, we derive a closed-form expression for the Cramer-Rao bound (CRB) associated with this estimation problem. Finally, we assess the statistical performance of the proposed estimators via Monte Carlo simulations, by comparing their mean-square error with the CRB.

**Index Terms**—Tensor Decomposition, Structured CPD, Cramer-Rao bound, Wiener-Hammerstein model

## 1. INTRODUCTION

The canonical polyadic decomposition (CPD), which can be seen as one possible extension of the SVD to higher-order tensors [1], is by now a well-established mathematical tool utilized in many scientific disciplines [2]. As it requires only mild assumptions for being essentially unique, the CPD provides means for blindly and jointly identifying the components of multilinear models, which arise in many real-world applications; see [1–3] for some examples.

In particular, the computation of CPDs having structured factors—such as Vandermonde, Toeplitz or Hankel matrices—has been shown useful in problems including channel estimation [4], nonlinear system identification [5] and multidimensional harmonic retrieval [6]. As a consequence, several special-purpose algorithms have been developed [6–9].

In practice, the data tensor to be decomposed is always corrupted by noise. Therefore, the assessment of the statistical performance of CPD computation algorithms via comparison with the Cramér-Rao bound (CRB) [10] is of practical interest, since it can guide the choice for an appropriate algorithm in application domains. Furthermore, it can provide valuable information for the study and development of such algorithms. For the unstructured CPD, [11] has derived the associated CRB and presented an evaluation of the popular alternating least-squares (ALS) algorithm for tensors of orders three and four. Regarding the structured case, the CRB for the estimation of a complex CPD with a particular Vandermonde factor has been derived in [12], motivated by the problem of estimating the directions of arrival of multiple source signals. Also, [13] has provided a closed-form expression for the CRB associated with the estimation of a CPD having Hankel and/or Toeplitz factors.

This paper addresses the statistical evaluation of algorithms specialized in computing a CPD having banded circulant factors when applied to estimate the parameters of a Wiener-Hammerstein (WH) model, which is a well-known block-oriented model used for representing nonlinear dynamical systems [14]. Because many systems of practical relevance can be (approximately) described by the WH model, the problem of identifying its parameters from a set of experimental data (*i.e.*, measured input and output samples) is well-studied; see, *e.g.*, [14] and references therein. One possible approach, as described in [5], consists in estimating the WH model parameters by computing the structured CPD of a kernel of its equivalent Volterra model. Here, we derive a closed-form expression for the CRB associated with this estimation problem, assuming the availability of a previously identified Volterra kernel corrupted by white Gaussian noise. Then, we formulate specialized estimation algorithms based on the CP Toeplitz (CPTOEP) and circulant-constrained ALS (CALS) methods proposed in [8, 9] and evaluate their performance by comparing their mean-square error with the CRB through Monte Carlo simulations.

The authors wish to thank DIGITEO for its financial support on the project DETMOTS-2A.

<sup>§</sup> Sponsored by CNPq–Brazil (individual grant 245358/2012-9).

<sup>‡</sup> Supported by ERC grant AdG-2013-320594 DECODA.

**Notation.** Scalars are denoted by lowercase letters, *e.g.*  $\theta_i$

or  $a_{ij}$ , vectors by lowercase boldface, *e.g.*  $\boldsymbol{\theta}$  or  $\mathbf{a}_j$ , matrices by boldface capitals, *e.g.*  $\mathbf{B}$  or  $\mathbf{A}^{(p)}$ , and higher order arrays by calligraphic letters, *e.g.*  $\mathcal{X}$ . We use the superscripts  $T$  for transposition,  $^\dagger$  for pseudo-inverse,  $\boxtimes$  and  $\odot$  denote the Kronecker and Khatri-Rao products, respectively, and  $\otimes$  stands for the (tensor) outer product. The shorthand  $\mathbf{a}^{\boxtimes p}$  denotes  $\mathbf{a} \boxtimes \dots \boxtimes \mathbf{a}$ , where  $\mathbf{a}$  appears  $p$  times;  $\mathbf{a}^{\otimes p}$  and  $\mathbf{A}^{\odot p}$  are defined analogously. For our purposes, a tensor  $\mathcal{X}$  of order  $P$  will be assimilated to its array of coordinates, which is indexed by  $P$  indices. Its entries will be denoted by  $x_{i_1, \dots, i_P}$ .

## 2. WIENER-HAMMERSTEIN MODEL IDENTIFICATION VIA CPD

### 2.1. Tensors and the CP decomposition

The polyadic decomposition of a  $p$ th-order tensor is defined by

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \otimes \dots \otimes \mathbf{a}_r^{(p)}, \quad (1)$$

where  $\mathbf{a}_r^{(q)}$  is the  $r^{\text{th}}$  column of  $\mathbf{A}^{(q)} \in \mathbb{R}^{I_q \times R}$ . The minimal value of  $R$  such that  $\mathcal{X}$  can be written as in (1) is called the rank of  $\mathcal{X}$ , in which case we refer to the above decomposition as the CPD of  $\mathcal{X}$ . Another way of expressing (1) is

$$\mathcal{X} = \mathcal{J} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_p \mathbf{A}^{(p)},$$

where  $\mathcal{J} \in \mathbb{R}^{R \times \dots \times R}$  is a  $p$ th-order diagonal tensor such that  $[\mathcal{J}]_{r, \dots, r} = 1$  and  $\times_q$  denotes the mode- $q$  product (see, *e.g.*, [2, Sec. 2.5]).

### 2.2. The WH model and its equivalent Volterra model

The structure of a discrete-time WH model is as depicted in Fig. 1. Basically, it consists of a cascade connection comprising a memoryless nonlinearity  $g(\cdot)$  “sandwiched” by two linear systems,  $W(z)$  and  $H(z)$ . Because of its structured form constituted by fundamental blocks, the WH model is said to belong to the class of block-oriented models [14].

In this paper, we consider the time-invariant WH model constituted by a polynomial nonlinearity  $g(x) = \sum_{p=1}^P g_p x^p$  and by finite impulse response filters  $W(z) = \sum_{l=0}^{L_w-1} w_l z^{-l}$ , with  $w_0 \neq 0$ , and  $H(z) = \sum_{r=0}^{R-1} h_r z^{-r}$ . Hence, the resulting expression relating the input  $u(n)$  to the output  $y(n)$  is

$$y(n) = \sum_{p=1}^P g_p \sum_{r=0}^{R-1} h_r \left[ \sum_{m=r}^{L_w+r-1} w_{m-r} u(n-m) \right]^p. \quad (2)$$

After some manipulation, this relation can be put in the equivalent Volterra model form

$$y(n) = \sum_{p=1}^P \sum_{m_1=0}^{M-1} \dots \sum_{m_p=0}^{M-1} k^{(p)}(m_1, \dots, m_p) \prod_{q=1}^p u(n-m_q),$$

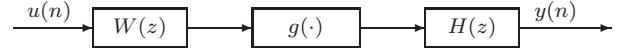


Fig. 1. Block-diagram of the Wiener-Hammerstein model.

whose symmetric discrete-time Volterra kernels are (uniquely) given by [15]

$$k^{(p)}(m_1, \dots, m_p) = g_p \sum_{l=l_0}^L h_l \prod_{q=1}^p w_{m_q-l}, \quad (3)$$

with  $M = L_w + R - 1$ ,  $l_0 = \max\{0, m_1 - L_w + 1, \dots, m_p - L_w + 1\}$  and  $L = \min\{R - 1, m_1, \dots, m_p\}$ .

### 2.3. CPD-based WH model identification

We now describe the WH model identification approach proposed in [5], which involves computing the CPD of a symmetric high-order Volterra kernel. We start by noting that, being a function of multiple discrete indices, any  $p$ th-order symmetric Volterra kernel  $k^{(p)}$  of memory  $M$  can be uniquely identified with a  $p$ th-order symmetric tensor  $\mathcal{X} \in \mathbb{R}^{M \times \dots \times M}$  defined by  $x_{m_1, \dots, m_p} = k^{(p)}(m_1 - 1, \dots, m_p - 1)$ . Owing to its convolutional form involving separable terms, the kernel in (3) can be identified with the tensor

$$\mathcal{X} = g_p \sum_{r=1}^R h_{r-1} \mathbf{c}_r^{\otimes p} = g_p \sum_{r=1}^R h_{r-1} (\mathbf{S}_r \mathbf{w})^{\otimes p}, \quad (4)$$

where  $\mathbf{S}_r \triangleq [\mathbf{e}_r \dots \mathbf{e}_{L_w+r-1}]$ , with  $\mathbf{e}_m$  denoting the  $m$ th canonical basis vector of  $\mathbb{R}^M$ , and  $\mathbf{w} = [w_0 \dots w_{L_w-1}]^T$ . Expression (4) is a symmetric CPD that can also be written as

$$\mathcal{X} = \mathcal{J} \times_1 \mathbf{C} \times_2 \dots \times_{p-1} \mathbf{C} \times_p [g_p \mathbf{C} \text{diag}(\mathbf{h})], \quad (5)$$

where  $\mathbf{h} = [h_0 \dots h_{R-1}]^T$  and  $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R] \in \mathbb{R}^{M \times R}$ . Note that the choice of which factor is postmultiplied by  $\text{diag}(\mathbf{h})$  is irrelevant, due to the scaling indeterminacy. We thus conclude that the WH model (2) has equivalent symmetric Volterra kernels whose CPD are constituted by circulant factors  $\mathbf{C}$  and a factor of the form  $g_p \mathbf{C} \text{diag}(\mathbf{h})$ , which absorbs the scaling coefficients  $g_p$  and  $h_r$ .

As the factors in (5) contain the parameters of the linear blocks of the WH model (2), the above observations suggest the following three-step procedure for its identification: (i) estimate  $k^{(p)}$  from an available set of input/output samples, using some Volterra kernel identification method (as, *e.g.*, [16]), (ii) compute the structured CPD from the associated symmetric tensor  $\mathcal{X}$  and (iii) estimate the coefficients  $g_q$ ,  $q \neq p$ , in the least-squares sense as explained in [5]. Note that this requires choosing some  $p \geq 3$ , since otherwise the model is not identifiable: for  $p = 1$ , it is a vector containing sums of products of coefficients  $g_p$ ,  $h_r$  and  $w_l$ ; for  $p = 2$ , we have a bilinear decomposition, which is only unique under restrictive assumptions (such as orthogonality). Henceforth, we assume that (i) has been accomplished and focus on step (ii).

### 3. ANALYTICAL CRB FOR CPD-BASED WH ESTIMATION ALGORITHMS

#### 3.1. Formulation of estimation problem

Let us consider that a  $p$ th-order tensor has been constructed from a non-null estimated kernel  $k^{(p)}$ , as described in the previous section. In practice, it is evident that such a tensor satisfies  $\mathcal{Y} = \mathcal{X} + \mathcal{N}$ , where  $\mathcal{N}$  is an error tensor accounting for the inevitable uncertainties which arise in the data-driven kernel estimation procedure. Furthermore, since  $k^{(p)}(m_1, \dots, m_p)$  is symmetric in  $m_1, \dots, m_p$ , in practice one estimates only the elements whose indices pertain to a suitable non-redundant domain such as  $D = \{(m_1, \dots, m_p) : m_1 \leq \dots \leq m_p\}$ , determining the others by symmetry. Hence,  $\mathcal{Y}$  and  $\mathcal{N}$  are also  $p$ th-order symmetric tensors, containing redundant elements. Introducing the selection matrix  $\Psi \in \mathbb{R}^{I \times M^p}$ , where  $I = |D| = \binom{M+p-1}{p}$ , which contains as rows<sup>1</sup> every product of the form  $\mathbf{e}_{m_p}^T \boxtimes \dots \boxtimes \mathbf{e}_{m_1}^T$  for  $(m_1, \dots, m_p) \in D$ , we can write the (non-redundant) vectorized model

$$\mathbf{y} \triangleq \Psi \text{vec}(\mathcal{Y}) = \mathbf{x} + \mathbf{n} \in \mathbb{R}^I,$$

where  $\mathbf{x} = \Psi \text{vec}(\mathcal{X})$  and  $\mathbf{n} = \Psi \text{vec}(\mathcal{N})$  is a random vector. Now, from (4), we can deduce

$$\begin{aligned} \text{vec}(\mathcal{X}) &= g_p \sum_{r=1}^R h_{r-1} (\mathbf{S}_r \mathbf{w})^{\boxtimes p} = \left[ g_p \sum_{r=1}^R h_{r-1} \mathbf{S}_r^{\boxtimes p} \right] \mathbf{w}^{\boxtimes p} \\ &= \left[ g_p \sum_{r=1}^R h_{r-1} \Phi_r \right] \mathbf{w}^{\boxtimes p} = \Phi(\mathbf{h}) \mathbf{f}(\mathbf{w}), \end{aligned} \quad (6)$$

where  $\Phi(\mathbf{h})$  is given by the term between brackets, in which  $\Phi_r = \mathbf{S}_r^{\boxtimes p}$ , and  $\mathbf{f}(\mathbf{w}) = \mathbf{w}^{\boxtimes p}$ .

Our problem can therefore be expressed as that of estimating the parameters  $g_p$ ,  $\mathbf{w}$  and  $\mathbf{h}$  of the WH model from observations which satisfy  $\mathbf{y} = \Psi \Phi(\mathbf{h}) \mathbf{f}(\mathbf{w}) + \mathbf{n}$ . We assume that the random vector  $\mathbf{n}$  has zero-mean i.i.d. components drawn from the Gaussian distribution with variance  $\sigma^2$ .

#### 3.2. Identifiability

Due to the inherent scaling indeterminacy of our model, its local identifiability is only guaranteed with further assumptions. To eliminate this indeterminacy, we assume  $w_0 = g_p = 1$ , which is sufficient due to the model structure. Note that this entails no loss of generality, as  $\mathbf{h}$  and the other coefficients  $w_l$  can be rescaled accordingly. Defining now  $\hat{\mathbf{w}}$  such that  $\mathbf{w} = [1 \ \hat{\mathbf{w}}^T]^T$ , we can write the parameter vector of the WH model as  $\boldsymbol{\eta} = [\hat{\mathbf{w}}^T \ \mathbf{h}^T]^T \in \mathbb{R}^M$ . Global identifiability, on the other hand, is related to the uniqueness of the structured CPD. As the  $k$ -rank [2, 3] of  $\mathbf{C}$  equals  $R$ , uniqueness follows from Kruskal's condition [2, Sec. 3.2] if  $\|\mathbf{h}\|_0 = R$  (which

implies that the  $k$ -rank of  $\mathbf{C} \text{diag}(\mathbf{h})$  is  $R$ ) and  $R \geq 2$ . If  $\|\mathbf{h}\|_0 < R$ , then the  $k$ -rank of  $\mathbf{C} \text{diag}(\mathbf{h})$  equals zero; in this case, Kruskal's condition is only met for  $P = 4$  if  $R \geq 3$  and for  $P \geq 5$  if  $R \geq 2$ .

#### 3.3. Parameter estimation algorithms

In this section, we briefly review two methods that can be used to estimate the parameters  $\boldsymbol{\eta}$  of a model of the form (4).

##### 3.3.1. Circulant-constrained ALS algorithm

The first method consists of a specialization of the well-known alternating least squares (ALS) algorithm in which the factor matrices of the CPD are constrained as in (5). In the case of a CPD involving only circulant factors, such strategy has already been followed in [9], leading to the CALS algorithm. Here, we adapt that algorithm for our purposes.

Initially, we define  $\mathbf{E}_l \triangleq [\mathbf{e}_l \ \dots \ \mathbf{e}_{R+l-1}] \in \mathbb{R}^{M \times R}$ , for  $l \in \{1, \dots, L_w\}$ , and  $\mathbf{E} \triangleq [\text{vec}(\mathbf{E}_1) \ \dots \ \text{vec}(\mathbf{E}_{L_w})]$ . With these definitions, we have  $\text{vec}(\mathbf{C}) = \mathbf{E} \mathbf{w}$ . Next, we note that any flat matrix unfolding of  $\mathcal{Y}$  can then be written as

$$\mathbf{Y} \approx \mathbf{C} \text{diag}(\mathbf{h}) (\mathbf{C}^{\odot p-1})^T = \text{unvec}_R(\mathbf{E} \mathbf{w}) \text{diag}(\mathbf{h}) (\mathbf{C}^{\odot p-1})^T$$

where the above approximation is due to the presence of noise and the operator  $\text{unvec}_R$  is defined such that,  $\forall \mathbf{a} = [\mathbf{a}_1^T \ \dots \ \mathbf{a}_R^T]^T$  with  $\mathbf{a}_r \in \mathbb{R}^N$ ,  $\text{unvec}_R(\mathbf{a}) = [\mathbf{a}_1 \ \dots \ \mathbf{a}_R]$ . Using the property  $\text{vec}(\mathbf{A} \text{diag}(\mathbf{b}) \mathbf{D}) = (\mathbf{D}^T \odot \mathbf{A}) \mathbf{b}$ , we have also  $\text{vec}(\mathbf{Y}) \approx (\mathbf{C}^{\odot p}) \mathbf{h}$ . Hence, given current estimates  $\hat{\mathbf{w}}^k$  and  $\hat{\mathbf{h}}^k$ , we can update them with the scheme

$$(i) \ \hat{\mathbf{v}}^{k+1} = \frac{1}{R} \mathbf{E}^T \text{vec} \left\{ \mathbf{Y} \left( \hat{\mathbf{W}}^k \right)^{\dagger} \left[ \text{diag}(\hat{\mathbf{h}}^k) \right]^{-1} \right\},$$

$$(ii) \ \hat{\mathbf{w}}^{k+1} = \frac{1}{[\hat{\mathbf{v}}^{k+1}]_1} \hat{\mathbf{v}}^{k+1},$$

$$(iii) \ \hat{\mathbf{h}}^{k+1} = \left( \hat{\mathbf{W}}^{k+1} \odot \hat{\mathbf{C}}^{k+1} \right)^{\dagger} \text{vec}(\mathbf{Y}),$$

where  $\hat{\mathbf{C}}^k = [\mathbf{S}_1 \hat{\mathbf{w}}^k \ \dots \ \mathbf{S}_R \hat{\mathbf{w}}^k]$  and  $\hat{\mathbf{W}}^k = (\hat{\mathbf{C}}^k)^{\odot p-1}$ . Note that, to derive (i), we have used  $\mathbf{E}^{\dagger} = (1/R) \mathbf{E}^T$ .

As stopping criteria, one can check whether the relative difference between two consecutive values of the reconstruction error  $J_{\mathbf{y}}^k = \left\| \mathcal{Y} - \mathcal{J} \times_1 \hat{\mathbf{C}}^k \times_2 \dots \times_p \hat{\mathbf{C}}^k \text{diag}(\hat{\mathbf{h}}^k) \right\|_F^2$  falls below some fixed threshold  $\epsilon_{\mathbf{y}} > 0$  or a maximum number of iterations  $K_{\max}$  is attained.

##### 3.3.2. CPTOEP algorithm

Since the objective is multimodal, the main goal is to find a good approximation of the solution by using a *low-complexity* algorithm. In [8], non-iterative procedures have been proposed, which are able to compute the *exact* CPD when matrix factors are banded or structured. Consider a matrix unfolding of  $\mathcal{Y}$  under the form:  $\tilde{\mathbf{Y}} \approx (\mathbf{C}^{(1)} \odot \mathbf{C}^{(2)}) \mathbf{A}^T$ , where the

<sup>1</sup>The ordering of the rows of  $\Psi$  is of no consequence for our purposes.

structure of  $\mathbf{A} = \mathbf{C}^{(3)} \odot \dots \odot \mathbf{C}^{(p)}$  is ignored, and where  $\mathbf{C}^{(n)}$  are assumed Toeplitz circulant of same size  $M \times R$ , that is, they can each be expressed in the orthonormal basis  $\{\mathbf{E}_\ell, 1 \leq \ell \leq L_w\}$  defined in Section 3.3.1:

$$\mathbf{C}^{(n)} = \sum_{\ell=1}^{L_w} c_\ell^{(n)} \mathbf{E}_\ell, \quad n \in \{1, \dots, p\}.$$

Let  $\tilde{\mathbf{Y}} = \mathbf{U}\Sigma\mathbf{V}^T$  denote the SVD of  $\tilde{\mathbf{Y}}$ . Then there exists a matrix  $\mathbf{N}$  such that  $\mathbf{UN} = (\mathbf{C}^{(1)} \odot \mathbf{C}^{(2)})$  and  $\mathbf{N}^{-1}\Sigma\mathbf{V}^T = \mathbf{A}^T$ . Following the lines of [8], one can find matrix  $\mathbf{N}$  and coefficients  $Z_{ij} = c_i^{(1)}c_j^{(2)}$  by solving a linear system of  $M^2R$  equations in  $L_w^2 + R^2 - 1$  unknowns. If there are more equations than unknowns and if the system has full rank  $R$ , the solution  $(\mathbf{N}, \mathbf{Z})$  is unique. First, coefficients  $c_i^{(1)}$  and  $c_j^{(2)}$  are obtained from the best rank-1 approximation of matrix  $\mathbf{Z}$ , which eventually yields estimates  $\hat{\mathbf{C}}^{(1)}$  and  $\hat{\mathbf{C}}^{(2)}$ . Next, we calculate  $\hat{\mathbf{C}} = (\hat{\mathbf{C}}^{(1)} + \hat{\mathbf{C}}^{(2)})/2$ , and the estimate of  $\mathbf{h}$  is obtained as in stage (iii) of the CALS algorithm.

The algorithm described above is suboptimal for several reasons: (a) the model is noisy, (b) the  $p$  factor matrices are assumed to be independent, whereas they are not, and (c) the structure of  $\mathbf{A}$  is ignored. Hence the solution obtained will be inaccurate, but can be easily refined by a quasi-Newton algorithm, as will be subsequently shown.

### 3.4. Closed-form expression for the CRB

If we assume that  $\boldsymbol{\eta}$  contains deterministic parameters associated with a system of interest, we have that the (vectorized) measured kernel satisfies  $\mathbf{y} \sim \mathcal{N}(\mathbf{x}, \sigma^2 \mathbf{I}_I)$ , where  $\sigma^2$  denotes the variance of the elements of  $\mathbf{n}$ . Hence, the mean-square error (MSE) of any locally unbiased estimator  $\hat{\boldsymbol{\eta}}(\mathbf{y})$  satisfies

$$E \left\{ \|\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}(\mathbf{y})\|^2 \right\} \geq \underbrace{\sum_{k=1}^{L_w-1} \text{CRB}(\tilde{w}_k) + \sum_{r=1}^R \text{CRB}(h_r)}_{\text{trace}(\mathbf{B}(\boldsymbol{\eta}))},$$

where the CRB matrix  $\mathbf{B}(\boldsymbol{\eta})$  can be computed by applying the Slepian-Bangs formula, which yields [13]

$$\mathbf{B}(\boldsymbol{\eta}) = \sigma^2 (\mathbf{J}(\boldsymbol{\eta})^T \mathbf{J}(\boldsymbol{\eta}))^{-1},$$

where  $\mathbf{J}(\boldsymbol{\eta}) \in \mathbb{R}^{I \times M}$  is the Jacobian matrix given by

$$\mathbf{J}(\boldsymbol{\eta}) = [\mathbf{J}(\tilde{\mathbf{w}}) \quad \mathbf{J}(\mathbf{h})] = \begin{bmatrix} \frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{w}}} & \frac{\partial \mathbf{x}}{\partial \mathbf{h}} \end{bmatrix}.$$

From (6) and the definition of  $\mathbf{f}$ , we have

$$\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{w}}} = \Psi \Phi(\mathbf{h}) \frac{\partial \mathbf{f}}{\partial \tilde{\mathbf{w}}} = \Psi \Phi(\mathbf{h}) [\mathbf{z}_1(\tilde{\mathbf{w}}) \quad \dots \quad \mathbf{z}_{L_w-1}(\tilde{\mathbf{w}})],$$

in which  $\mathbf{z}_l(\tilde{\mathbf{w}}) = \sum_{q=1}^p \mathbf{w}^{\boxtimes q-1} \boxtimes \mathbf{e}_{l+1} \boxtimes \mathbf{w}^{\boxtimes p-q}$  (with the convention  $\mathbf{w}^{\boxtimes 0} = 1$ ). To derive  $\mathbf{J}(\mathbf{h})$ , we first apply the property  $\text{vec}(\mathbf{ABD}) = (\mathbf{D}^T \boxtimes \mathbf{A})\text{vec}(\mathbf{B})$  to write

$$\mathbf{x} = \text{vec}(\Psi \Phi(\mathbf{h}) \mathbf{f}(\mathbf{w})) = (\mathbf{f}^T(\mathbf{w}) \boxtimes \Psi) \text{vec}(\Phi(\mathbf{h})),$$

leading thus to

$$\frac{\partial \mathbf{x}}{\partial \mathbf{h}} = (\mathbf{f}^T(\mathbf{w}) \boxtimes \Psi) [\text{vec}(\Phi_1) \quad \dots \quad \text{vec}(\Phi_R)].$$

In order to identify the contribution of  $\mathbf{w}$  and  $\mathbf{h}$  in  $\text{CRB}(\tilde{w}_k)$  and  $\text{CRB}(h_r)$ , we propose to extend the results presented in [13] by using oblique projection. This is the purpose of the following proposition. We denote by  $\mathbf{E}_{\mathbf{AB}}$  the oblique projection whose range is  $\langle \mathbf{A} \rangle$  and whose null space contains  $\langle \mathbf{B} \rangle$  (see [17] for details).

**Proposition 3.1** *The closed-form expression for the CRB of  $\tilde{w}_k$  is given by:*

$$\text{CRB}(\tilde{w}_k) = \frac{\sigma^2}{\|\mathbf{g}_k\|^2 - \|\mathbf{E}_{\mathbf{G}_k \mathbf{J}(\tilde{\mathbf{w}})} \mathbf{g}_k\|^2 - \|\mathbf{E}_{\mathbf{J}(\tilde{\mathbf{w}}) \mathbf{G}_k} \mathbf{g}_k\|^2},$$

where  $\mathbf{g}_k$  is the  $k$ th column of  $\mathbf{J}(\tilde{\mathbf{w}})$  and  $\mathbf{G}_k$  is the submatrix of  $\mathbf{J}(\tilde{\mathbf{w}})$  obtained by removing its  $k$ th column. Similarly, the closed-form expression for the CRB of  $h_r$  is:

$$\text{CRB}(h_r) = \frac{\sigma^2}{\|\mathbf{d}_r\|^2 - \|\mathbf{E}_{\mathbf{D}_r \mathbf{J}(\tilde{\mathbf{w}})} \mathbf{d}_r\|^2 - \|\mathbf{E}_{\mathbf{J}(\tilde{\mathbf{w}}) \mathbf{D}_r} \mathbf{d}_r\|^2},$$

where  $\mathbf{d}_r$  is the  $r$ th column of  $\mathbf{J}(\mathbf{h})$  and  $\mathbf{D}_r$  is the submatrix of  $\mathbf{J}(\mathbf{h})$  obtained by removing its  $r$ th column.

The proof is omitted due to the lack of space.

## 4. SIMULATION RESULTS

To illustrate the utility of the derived CRB, we now present some Monte Carlo simulation results. Specifically, we evaluate several estimators when applied to identify a WH model with parameters  $\mathbf{w}^T = [1 \ 0.538 \ 1.834 \ -2.259 \ 0.862]^T$ ,  $\mathbf{h} = [1.594 \ -6.538 \ -2.168]^T$  from estimates of the equivalent symmetric third-order kernel  $\mathcal{X}$ , proceeding as follows. For each realization of the (symmetric) noise tensor  $\mathcal{N}$ , we vary  $\sigma^2$  and then construct a data tensor  $\mathcal{Y} = \mathcal{X} + \mathcal{N}$  for each chosen level of  $\sigma^2$ . Next, we compute estimates  $\hat{\boldsymbol{\eta}}(\mathbf{y}) \in \mathbb{R}^7$  given by: (i) the family of estimators  $N$ -CALS, which consist in applying  $N$  times the algorithm of Section 3.3.1 with random initializations and keeping the best solution in terms of reconstruction error (w.r.t.  $\mathcal{Y}$ ); (ii) the estimator CPTOEP, described in Section 3.3.2; (iii) the estimator CPTOEP-CALS, which corresponds to refining the CPTOEP estimate by applying the CALS algorithm; (iv) the estimator CPTOEP-BFGS, in which a similar refinement is obtained by minimizing a least-squares criterion (w.r.t.  $\mathcal{Y}$ ) with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm<sup>2</sup> [18]. The maximum number of iterations established for CALS and BFGS is  $K_{\max} = 2000$ . We choose  $\epsilon_{\mathcal{Y}} = 10^{-10}$  and set the tolerance of BFGS also as  $10^{-10}$ . For each estimate  $\hat{\boldsymbol{\eta}}(\mathbf{y})$ , we compute  $\epsilon_{\boldsymbol{\eta}} = \|\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}(\mathbf{y})\|^2$ . This procedure is repeated for 100 realizations of  $\mathcal{N}$  and then  $\epsilon_{\boldsymbol{\eta}}$  is averaged for each level of  $\sigma^2$ , yielding a mean-square error estimate denoted by  $\text{MSE}_{\boldsymbol{\eta}}$ .

<sup>2</sup>Specifically, we used the Fortran implementation whose Matlab interface is available at <http://github.com/pcarbo/lbfgsb-matlab>.



**Table 1.** Simulation results: estimated  $\text{MSE}_{\eta}$  values (in dB).

Estimator	$1/\sigma^2$ (dB)					
	10	20	30	40	50	60
1-CALS	19.22	17.14	18.37	17.68	18.53	17.86
5-CALS	-15.04	-25.05	4.04	4.05	-55.07	4.06
10-CALS	-15.04	-25.05	-35.04	-45.07	-55.02	-65.07
CPTOEP	-13.96	-23.94	-33.94	-43.94	-53.94	-63.94
CPTOEP-CALS	-15.04	-25.04	-35.04	-45.05	-55.02	-65.13
CPTOEP-BFGS	-20.04	-30.03	-40.03	-50.02	-60.01	-69.62
CRB	-20.18	-30.18	-40.18	-50.18	-60.18	-70.18

The results are shown in Table 1, as well as the computed values of the CRB. One can see that 1-CALS has a very poor performance, due to its frequent premature termination or inability to converge. Although 5-CALS performs better, its results are degraded for the same reasons. CPTOEP, in its turn, performs slightly worse than 10-CALS, but attains a similar level when refined by CALS. Yet, there remains a gap between their MSE curves and that of the CRB. Indeed, only CPTOEP-BGFS attains an MSE close to the CRB. Note that a similar gap has been reported by [11] for the ALS algorithm. Along the lines of their discussion, we believe that, in the case of CALS, this gap is due to the convergence problems which are always observed in practice, at least for a few runs. As for CPTOEP, this seems to happen because the adapted procedure yields suboptimal estimates.

Finally, we note that the above comparison is justified since, under the assumption of Gaussian additive noise, the least-squares criterion leads to the maximum likelihood (ML) estimator. In signal-in-noise problems, the ML estimator is often approximately unbiased even for a small sample size, provided that the SNR is sufficiently high [10].

## 5. CONCLUSION

A closed-form expression of the CRB has been derived for the parameter estimates of a CPD having identical banded circulant factors, one of which is post-multiplied by a diagonal scaling matrix. Then, two specialized algorithms have been proposed to compute a CPD with that structure. The first, named CALS, is an adaptation of the ALS method taking the structural constraints into account, whereas the second is composed of two steps: (i) compute an approximate solution thanks to a non iterative algorithm (CPTOEP), and (ii) refine the solution via CALS or via a quasi-Newton descent (BFGS). The latter (CPTOEP-BFGS) reached the Cramér-Rao bound over a wide range of SNR values. The proposed algorithms have been applied to identify a WH model, and their statistical performance has been evaluated using the derived CRB.

## REFERENCES

[1] P. Comon, "Tensors : A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, 2014.

[2] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[3] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, 2014.

[4] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "Blind channel identification algorithms based on the PARAFAC decomposition of cumulant tensors: the single and multiuser cases," *Signal Process.*, vol. 88, no. 6, pp. 1382–1401, 2008.

[5] G. Favier and A. Y. Kibangou, "Tensor-based methods for system identification. Part 2: Three examples of tensor-based system identification methods," *Int. J. Sci. Tech. Automat. Control (IJ-STA)*, vol. 3, no. 1, pp. 870–889, 2009.

[6] M. Sorensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, 2013.

[7] A. Y. Kibangou and G. Favier, "Non-iterative solution for PARAFAC with a Toeplitz matrix factor," in *Proc. EUSIPCO*, Glasgow, UK, 2009, pp. 691–695.

[8] M. Sorensen and P. Comon, "Tensor decompositions with banded matrix factors," *Linear Alg. Appl.*, vol. 438, no. 2, pp. 919–941, 2013.

[9] J. H. de M. Goulart and G. Favier, "An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 4, pp. 1543–1562, 2014.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Fundamentals of Statistical Signal Processing. PTR Prentice-Hall, 1993.

[11] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, 2001.

[12] S. Sahnoun and P. Comon, "Tensor polyadic decomposition for antenna array processing," in *Proc. Int. Conf. Comput. Stat. (Compstat)*, Geneva, Switzerland, 2014, pp. 233–240.

[13] M. Boizard, R. Boyer, G. Favier, and P. Comon, "Performance estimation for tensor CP decomposition with structured factors," in *Proc. ICASSP*, 2015.

[14] R. Haber and L. Keviczky, *Nonlinear system identification – Input-Output Modeling Approach*, vol. 1 of *Mathematical Modelling*, Kluwer Academic Publishers, 1999.

[15] A. Y. Kibangou and G. Favier, "Wiener-Hammerstein systems modeling using diagonal Volterra kernels coefficients," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 381–384, 2006.

[16] A. Y. Kibangou and G. Favier, "Identification of fifth-order Volterra systems using i.i.d. inputs," *IET signal process.*, vol. 4, no. 1, pp. 30–44, 2010.

[17] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1413–1424, 1994.

[18] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.